# **1. Линейный слой**

# **Что такое линейный (полносвязный) слой? Напишите формулу. Сформулируйте Универсальную теорему аппроксимации. Зачем нужна нелинейность между линейными слоями? Какие нелинейности вам известны? В чем их плюсы и минусы? Как вычисляется число обучаемых параметров в линейном слое? Как работает backpropagation? Покажите на примере маленькой полносвязной нейросети, расписав формулы.**

*Линейный (полносвязный) слой* — это слой нейронной сети, где каждый нейрон соединён со всеми входами предыдущего слоя.

**y**=**Wx**+**b**,

где:

* – входной вектор,
* **W** – матрица весов,
* **b** – вектор смещений,
* **y** – выход слоя.

Всего параметров m(n+1), где n – кол-во входов, m – кол-во выходов.

*Универсальная теорема аппроксимации*

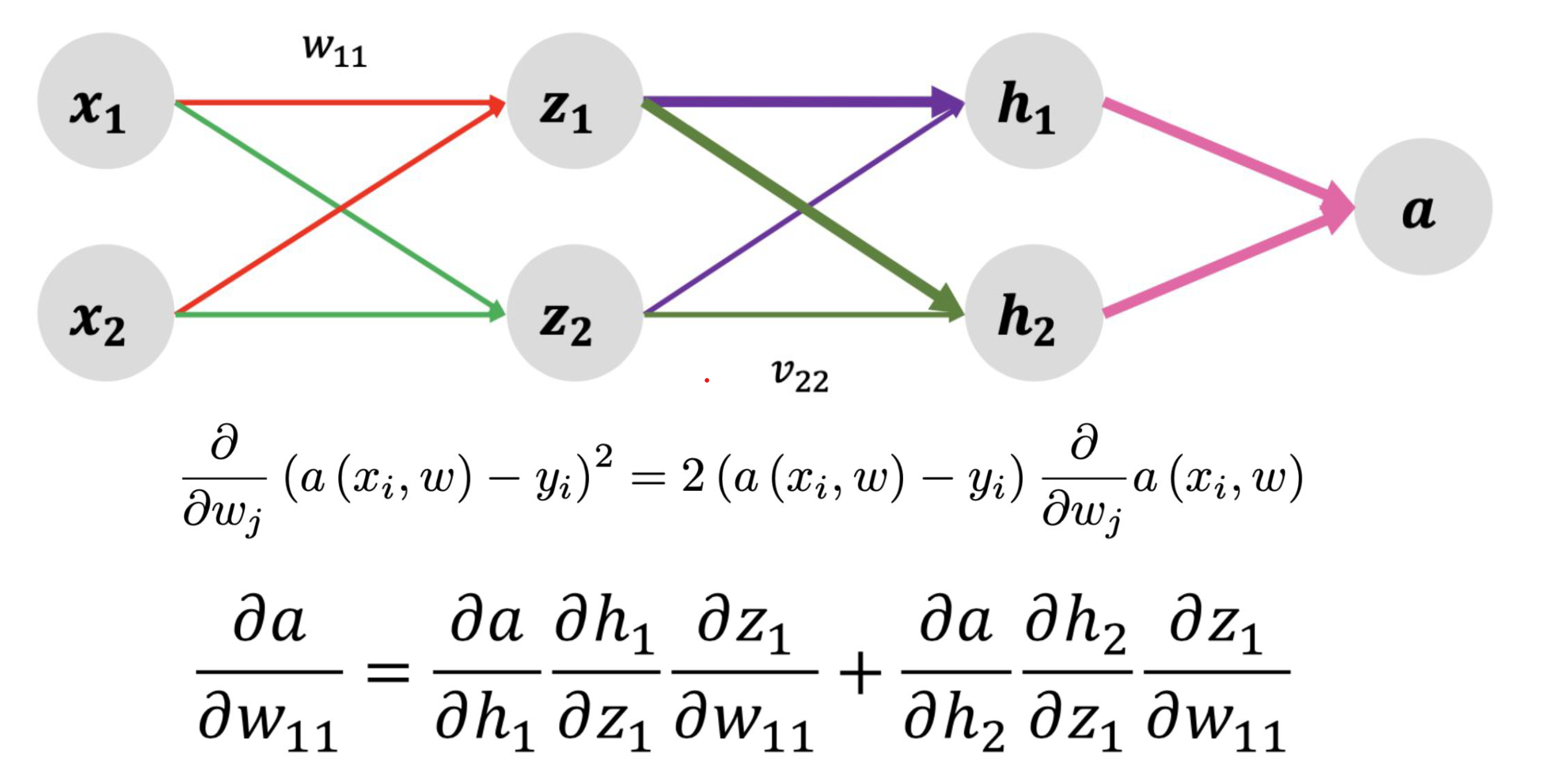
Имея некую непрерывную функцию f(x), мы можем построить такую двухслойную нейронную сеть, что будем приближать f(x) с любой заранее заданной точностью.

*Нелинейность*

Без нелинейностей композиция линейных слоев эквивалентна **одному линейному преобразованию**:

|  |  |  |  |
| --- | --- | --- | --- |
| Функция | Формула | Плюсы | Минусы |
| **ReLU** | max(0,*x*) | Быстрая сходимость, простота | "Умирающие нейроны" при x<0 |
| **Сигмоида** |  | Гладкая, выход в [0,1] | Затухание градиента при увеличении x |
| **Гиперболический тангенс** |  | Симметрична (выход [−1,1]) | Те же проблемы, что у сигмоиды |

Как работает backpropagation?



# **2. Градиентный спуск и loss-функции**

# **Опишите основные виды градиентного спуска (GD, SGD, mini-batch). Что такое Adam и AdamW? Какие loss-функции применяются для задач классификации и регрессии? Как влияет выбор lr и использование lr-scheduler на процесс обучения?**